

David Schreiber, Cerstin Mahlow and Britta Juska-Bacher

Phraseologische Neologismen: Identifikation und Validierung¹

Abstract: The study described in this article investigates the identification of phraseological neologisms in German based on a combination of methods and resources.

Detecting phraseological neologisms is a difficult task. In general, detecting phrasemes in corpora is hampered by their low number of occurrences. Most corpora contain written language; there are almost no big corpora of spoken German suitable for phraseme detection. However, phraseological neologisms are expected to first occur in spoken language before being encountered in written texts (Burger 2010).

Based on a general definition of neologisms, a definition of phraseological neologisms is suggested. A phraseme is considered as neologism (1) if it is a new form with a meaning not known previously or (2) if a known form has a new phraseological meaning from a certain point in time. A neologism has to be known by a reasonable number of speakers in a reasonably large region. For practical reasons, in this study German phraseological neologisms are defined as occurring after the beginning of the year 2000.

According to Bergenholz and Mugdan (1991), lexicographers might use various methods to study lexical units: harvesting existing collections, asking informants, collecting evidence, searching corpora, or using introspection. Except for relying on their own intuition, the other methods are useful to some extent. To achieve best results, a combination of all methods for identifying potential neologisms as well as for validation is proposed.

The reported study is a proof of concept of the proposed combination of methods, and investigates how to identify and validate phraseological neologisms given two current collections of phrasemes: a printed one, published in 2008, and a web-based one, updated and expanded regularly. Since the online collection is constantly growing, it can be assumed that new entries might include neologisms. The printed collection is assumed to contain established phrasemes, since the underlying resources are mainly corpora of written German (with a main focus on the post-war period). Moreover, the process of phraseme

¹ Gedankt sei den zwei anonymen Referent(inn)en und den Herausgeber(innen) des *Yearbook of Phraseology* für ihre konstruktiven Kritikpunkte zu einer früheren Version dieses Artikels.

collection had to be finished before the publication date, i.e., 2008, which means that no phrasemes that emerged after 2008 could be included.

Starting with the automatic extraction of unique phrasemes from both collections (13,287 from the printed collection and 12,235 from the online collection), a pragmatic comparison is carried out with the help of natural language processing resources. The collections follow different guidelines for formulating phrasemes, making it impossible to automatically compare the entries as such. Sometimes variants for verbs or nouns are mentioned explicitly, fillers are sometimes given as an impersonal form like *jemand* ('someone') or *jemanden* ('someone', Dative Singular) – both in the full form or as an abbreviation – sometimes as an example, or as a sequence of possible fillers.

To overcome this situation, all entries were processed by an automatic morphological analyzer, resulting in a list of base forms (lemmas) of nouns and verbs used in the phrasemes. For each lemma the phrasemes in which it occurs were listed. The list of lemmas from the printed collection and the list of lemmas from the online collection were then compared resulting in 2,066 lemmas occurring in the online collection only. The 2,577 phrasemes these lemmas occur in are assumed to be potential phraseological neologisms, and not listed in the printed collection.

To validate the neologism candidates, several steps were carried out. First, all one-word items were removed as they did not match the requirement of polylexicity, as defined by Burger (2010).

Second, manual effort was needed to identify phrasemes that were only variants of phrasemes already listed in the printed collection like *sich kringelig lachen* (literally 'to laugh yourself curly') which is a variant of *sich kringeln vor Lachen/Vergnügen* (literally 'to curl because of laughter/joy') in the printed collection. Sometimes this was also due to different guidelines for formulating phrasemes as for *böses Blut machen/schaffen* (literally 'to make/produce bad blood') in the printed collection vs. *böses Blut erzeugen* (literally 'to create bad blood') in the list of potential neologisms originating from the online collection. This reduced the list of potential neologisms by half, resulting in 1006 remaining candidates.

The third step was a manual lookup of the neologism candidates in several collections printed in the 21st century and in available online corpora. This resulted in a reduced list of 79 remaining candidates.

Up to this step it is possible to state that there is no evidence for the existence of these 79 phrasemes in accessible collections and corpora. However, most of the collections are based on written corpora and most of the corpora consulted contain written language. However, it would be possible that some phrasemes have been known for a long time, but occur in spoken language only.

To test this, a questionnaire was sent out to respondents from all German-speaking regions in Germany, Austria, and Switzerland. The interpretation of the

data resulted in only 15 phrasemes which were known by more than 20% of the informants and categorized by at least one third of them as neologism (not known before year 2000). Some of the phrasemes were known to all respondents, some phrasemes were unknown to all respondents.

The resulting list of German phraseological neologisms might not be fully correct, but the reported study shows that a clever combination of methods for detecting neologisms in general is a valuable approach to identify phraseological neologisms. This study also shows that online dictionaries or online collections serve as a good starting point to discover new language units and new meanings.

Keywords: neologism; lexicography; natural language processing; survey; combination of methods.

Correspondence address: Britta.Juska-Bacher@unibas.ch

1 Phraseologische Neologismen

Neologismen werden in der Phraseologieforschung bisher stiefmütterlich behandelt. Dies ist insbesondere auf methodische Schwierigkeiten bei der Identifikation dieser Einheiten zurückzuführen. Selbst etablierte Phraseme kommen in großen Korpora relativ selten vor (Colson 2007: 1072). Berücksichtigt man ausschließlich aktuelle Texte, reduziert sich die Größe solcher Korpora erheblich. Sie reichen dann nicht mehr aus, um Phraseme durch Kookkurrenzanalysen zu ermitteln. Gehören die Neologismen zudem in den Bereich der mündlichen Sprache – was, wie Burger (2010: 131; siehe auch Ehrhardt 2007: 258f.) vermutet, eher häufig der Fall ist – ist eine automatische Ermittlung in Korpora nahezu unmöglich. Im Deutschen existiert bislang kein Korpus der mündlichen Sprache, das groß genug und zudem aktuell wäre, um phraseologische Neologismen durch korpusgesteuerte Analysen suchen zu können.

Die Definition phraseologischer Neologismen (nach Herberg et al. 2004: XI: “Neuphraseologismen”) soll sich an diejenige von monolexikalischen Neologismen anlehnen, die üblicherweise in Abgrenzung zu gebräuchlichen Wortschatzeinheiten definiert, etwa:

Von den etablierten, seit langem gebräuchlichen Wortschatzeinheiten unterscheidet sich der Neologismus dadurch, dass 1. entweder seine Form **und** seine Bedeutung oder 2. **nur** seine Bedeutung von der Mehrheit der Angehörigen einer bestimmten Kommunikationsgemeinschaft von einem bestimmten (allerdings schwer bestimmbaren) Zeitpunkt an und über eine bestimmte Zeitphase hinweg als neu empfunden wird.

(Herberg/Kinne 1998: 1f)

Entsprechend sollen phraseologische Neologismen, die in dieser Arbeit das Merkmal [+ idiomatisch] tragen, wie folgt definiert werden:

Ein phraseologischer Neologismus unterscheidet sich von einem etablierten Phrasem dadurch, dass entweder 1. die Form **und** die phraseologische Bedeutung oder 2. **nur** die phraseologische Bedeutung der Mehrworteinheit ab dem Zeitpunkt *t* in der Sprachgemeinschaft überregional gebräuchlich ist und während eines bestimmten Zeitraums als neu empfunden wird.

Für phraseologische Neologismen wie für etablierte Phraseme gelten die Charakteristika Polylexikalität, Festigkeit und (fakultativ) Idiomatizität (Burger 2010: 15–32). Daher kann eine neue Mehrwortverbindung, die nicht nur eine Variante eines bereits bestehenden Phrasems sein darf, entweder mit einer phraseologischen Bedeutung verbunden werden (siehe 1.) oder aber eine bereits bestehende Mehrworteinheit erhält eine neue phraseologische Bedeutung (siehe 2.). In der Regel betrifft dies nicht-idiomatische Einheiten, die neu eine idiomatische Bedeutung erhalten.

Empirische Studien haben gezeigt, dass Phraseme einen sehr unterschiedlichen Bekanntheitsgrad besitzen, der auch bei den bekannteren Einheiten deutlich unter 50% liegen kann (Hallsteinsdóttir et al. 2006). Statt von einer Bekanntheit bei der “Mehrheit der Angehörigen einer bestimmten Kommunikationsgemeinschaft” (siehe die Definition von Herberg/Kinne 1998) auszugehen, beschränkt sich unsere Definition phraseologischer Neologismen daher auf eine überregionale Gebräuchlichkeit, um das Kriterium der Festigkeit in der Sprachgemeinschaft zu erfüllen. Wir verzichten explizit darauf, einen allgemein gültigen Mindestbekanntheitsgrad festzulegen. Dieser ist in Abhängigkeit vom phraseologischen Material individuell anzusetzen. Bei der Auswertung unserer empirischen Daten in Abschnitt 4.2 verlangen wir einen Bekanntheitsgrad von 20%.

Neologismen dürfen vor dem Zeitpunkt *t* mit dieser Bedeutung nicht belegt sein. Für diese Studie wird der Zeitpunkt *t* auf den 1.1.2000 festgelegt, da phraseologische Neologismen seither nicht systematisch untersucht wurden. Für die 1990er Jahre hingegen existiert beispielsweise das Neologismenwörterbuch “Neuer Wortschatz. Neologismen der 90er Jahre im Deutschen” (Herberg et al. 2004), das neben monolexikalischen auch phraseologische Neologismen aufführt.

Die Grundlage des vorliegenden Beitrags bildet der elektronisch gestützte Abgleich zweier idiomatischer Sammlungen zur Identifizierung potentieller phraseologischer Neologismen entsprechend der o.g. Definition. Als Quellen werden der *redensarten-index.de* und der *Duden Band 11 Redewendungen* verwendet. Der Abgleich basiert auf der Annahme, dass der *redensarten-index.de* als Online-Sammlung grundsätzlich die besseren Voraussetzungen hat,

Neologismen früh – wenn auch nicht systematisch – lexikographisch zu erfassen (siehe Abschnitt 3.1).

Diejenigen Einträge des *redensarten-index.de*, die nicht im *Duden 11* aufgeführt sind, werden automatisch extrahiert (siehe Abschnitt 3.2). Nach einer manuellen Bereinigung dieser Daten um Einheiten, die nicht der vorgegebenen Phrasemdefinition entsprechen (siehe Abschnitt 3.3), werden die ermittelten Mehrworteinheiten in einem zweiten Schritt anhand idiomatischer Sammlungen, Wörterbücher und Textkorpora auf ihr Vorkommen vor dem gegebenen Zeitpunkt *t* (der 1.1.2000) überprüft und ggf. aus der Liste gestrichen (siehe Abschnitt. 4.1). Um schließlich die Bekanntheit und Verbreitung im deutschsprachigen Raum sowie das Alter der potentiellen phraselogischen Neologismen zu überprüfen, werden in einem dritten Schritt Informanten in Deutschland, Österreich und der Schweiz zu den betreffenden Phrasemen befragt (siehe Abschnitt 4.2).

Wir beschränken uns hier auf deutschsprachige (teil-)idiomatische Phraseme. Typische nähe- oder distanzsprachliche Phänomene wie formelhafte Interaktions- oder Textorganisationssignale, von denen eine spezifische Verteilung auf gesprochene und geschriebene Sprache zu erwarten ist (Stein 2007), bleiben unberücksichtigt.

Ziel dieser Studie ist es, verschiedene Methoden und Ressourcen, insbesondere die Möglichkeiten einer Online-Sammlung, auf ihre Eignung zum Auffinden und Validieren phraselogischer Neologismen zu testen.

2 Methoden zum Umgang mit phraselogischen Neologismen

Für das Aufspüren phraselogischer Neologismen stehen uns etablierte empirische Erhebungsmethoden zur Verfügung, die bereits Bergenholtz/Mugdan (1991: 1612–1622) für die Lexikographie im Allgemeinen nennen:

- Auswertung vorhandener Beschreibungen, d.h. Forschungsliteratur, Wörterbücher, Sammlungen
- Informantenbefragungen
- Belegsammlungen
- Korpusanalysen

Die Introspektion der Forschenden, ebenfalls von Bergenholtz/Mugdan (1991) genannt, ist bei der Suche nach Neologismen weniger hilfreich. Der

Selbstbeobachtung mangelt es einerseits an Objektivität, andererseits wird sie nur eine sehr begrenzte Zahl an Neologismen liefern. Im Folgenden wird auf die verschiedenen Methoden genauer eingegangen, zunächst unter dem Aspekt der Ermittlung potentieller phraseologischer Neologismen (Abschnitt 2.1), anschließend unter dem Aspekt der Validierung (Abschnitt 2.2). Abschließend wird der hier gewählte Ansatz der Methodenkombination vorgestellt (Abschnitt 2.3).

2.1 Identifikation potentieller phraseologischer Neologismen

Auswertung vorhandener Beschreibungen: Sollen existierende Ressourcen ausgewertet werden, müssen diese vor relativ kurzer Zeit entstanden sind, d.h. nach der o.g. Definition von phraseologischen Neologismen ab dem Jahr 2000. Aus dem verfügbaren Material boten sich für unsere Zwecke der *Duden 11* und die Online-Sammlung *redensarten-index.de* an. Unseres Wissens liegt keine aktuelle Forschungsliteratur vor, die deutsche phraseologische Neologismen dieses Zeitraums in größerem Umfang anführt. Um mit Hilfe von aktuellen Wörterbüchern oder idiomatischen Sammlungen Neologismen zu identifizieren, ist ein Abgleich mit älteren Sammlungen nötig. Phraseme, die in älteren Werken nicht enthalten sind, können als potentielle Neologismen betrachtet und müssen weiter validiert werden (siehe Abschnitt 3).

Informantenbefragungen: Um phraseologische Neologismen mit Hilfe von Informanten zu ermitteln, müssen die Teilnehmenden in offener Form befragt werden, welche Phraseme ihrer Meinung nach innerhalb eines bestimmten Zeitraums, in unserem Fall der letzten 12 Jahre, entstanden sind. Unsere Untersuchung hat allerdings gezeigt, dass die Rückmeldungen eher spärlich ausfallen und sich diese Methode zur Ermittlung phraseologischer Neologismen weniger eignet (siehe Abschnitt 4.2). Deutlich wertvollere Dienste können Informanten bei der Validierung von potentiellen Neologismen liefern (siehe Abschnitt 2.2).

Belegsammlungen (der Forschenden): Wie besonders in der historischen Lexikographie üblich, können Belege potentieller Neologismen gesammelt werden. Mit dieser Methode sind im Fall von Neologismen jedoch drei Schwierigkeiten verbunden: (1) Neologismen können nicht systematisch, sondern nur punktuell erfasst werden, (2) die Einschätzung des Alters von Phrasemen ist – wie unsere Ergebnisse in Abschnitt 4.2.1 zeigen – sehr subjektiv und schwierig (siehe auch Herberg et al. 2004) und (3) diese Belege entsprechen vom Alter her bereits nach relativ kurzer Zeit nicht mehr der Neologismus-Definition. In jedem Fall müssen auf diese Weise klassifizierte potentielle Neologismen einer Validierung unterzogen werden.

Korpusanalysen: Phraseme sind aufgrund ihres seltenen Vorkommens (Colson 2007: 1072), ihrer prinzipiell möglichen wörtlichen wie phraseologischen Verwendung und ihrer Variabilität, die gerade bei einem sich noch im Verfestigungsprozess befindlichen neuen Phrasem besonders groß ist, in Korpora deutlich schwieriger zu suchen als ein Einzelwort:

Die Aufgabenstellung, das Web oder andere Texte nach neuen, bisher nicht dokumentierten Phraseologismen zu durchsuchen, ist auf den ersten Blick faszinierend, aber alles andere als einfach.

(Quasthoff et al. 2010: 51)

Bei der Suche nach Neologismen müssen die Texte des Korpus sehr neu sein (in unserem Beispiel ab dem Jahr 2000 entstanden). Wenn zudem erwartet wird, dass Neologismen aus der gesprochenen (Umgangs-)Sprache in die geschriebene übernommen werden (siehe Burger 2010: 131; Ehrhardt 2007: 258), sollte das Korpus einen möglichst großen Anteil gesprochener Sprache enthalten. Ein solches Korpus existiert bisher für das Deutsche nicht.

Eine Korpusanalyse der letzten 12 Jahre im Analyseparadigma (Begriff nach Steyer 2004: 93), d.h. mittels Kookkurrenzanalysen, um phraseologische Neologismen aufzuspüren, ist damit wenig erfolgversprechend. Hingegen ist die Nutzung von Korpora im Konsultationsparadigma (ebd.), d.h. die Suche nach einer vorgegebenen Liste von Phrasemen, zur Validierung potentieller Neologismen möglich.

2.2 Validierung phraseologischer Neologismen

Werden Phraseme als potentielle phraseologische Neologismen durch eine oder mehrere der oben beschriebenen Methoden identifiziert, müssen sie validiert werden. Dafür kommen prinzipiell die gleichen Methoden in Frage, sie eignen sich jedoch unterschiedlich gut.

Auswertung vorhandener Beschreibungen: Potentielle Neologismen werden mit älteren Wörterbüchern und Sammlungen abgeglichen. Einheiten, die vor dem Zeitpunkt *t* belegt sind, sind keine Neologismen. Wenn die betreffenden Phraseme nicht enthalten sind, bedeutet dies jedoch nicht notwendigerweise, dass es sich tatsächlich um Neologismen handelt. Phraseme können aus verschiedenen Gründen nicht in einer Sammlung enthalten sein: z.B. weil sie den Lexikographen unbekannt sind, weil sie in einer nicht fokussierten Stilschicht verwendet werden oder weil sie nur in der mündlichen Sprache Verwendung finden.

Informantenbefragungen: Anhand von Befragungen kann eruiert werden, in welcher Region des Sprachgebiets welche Phraseme wann und wem bekannt sind. Geht man davon aus, dass ein nicht unerheblicher Teil von Neologismen aus der Jugendsprache stammt (siehe z.B. Burger 2010: 131 oder Ehrhardt 2007: 259, der auf das besondere “Innovationspotential von [. . .] jugendsprachlichen Phrasemen” und ihre “Sprachwandel-Dynamik” hinweist), ist besonders der Faktor Alter relevant. Wenn es sich tatsächlich um Neologismen handelt, sollten sie eine definierte Mindestbekanntheit unter den Informanten (evtl. besonders unter den jüngeren von ihnen) haben und von einem möglichst großen Anteil der Informanten als “neu” eingestuft werden. Die Schätzung des Alters der Phraseme ist aber mit Schwierigkeiten verbunden, wie unsere Studie zeigt. Für einige Faktoren (wie die areale Verbreitung von Phrasemen und ihre Bekanntheit nach Alter der Probanden) stellt eine Befragung die einzig mögliche Erhebungsmethode dar.

Belegsammlungen: Die von den Forschenden gesammelten Belege eignen sich nicht für eine Validierung, da es sich lediglich um subjektive Einschätzungen Einzelner handelt, die einer empirischen Prüfung unterzogen werden sollten.

Korpora: Große Textsammlungen eignen sich sehr gut, um im Konsultationsparadigma abgefragt zu werden – sowohl für die Suche nach phraseologischen wie auch nach monolexikalischen Neologismen. Finden sich potentielle phraseologische Neologismen in Texten, die vor 2000 entstanden sind, entsprechen sie nicht unserer Definition von Neologismen. Sind die Phraseme nicht zu finden, ist dies kein zwingender Hinweis darauf, dass es sich um Neologismen handelt, besonders wegen des Fehlens geeigneter Korpora mit mündlichen Texten.

2.3 Methodenkombination

Zusammenfassend lässt sich feststellen, dass der Einsatz einzelner Methoden zur Ermittlung von phraseologischen Neologismen und deren Validierung nicht genügt. Ihre Verwendung ist notwendig, jedoch nicht hinreichend. Mit Hilfe einer der genannten Methoden kann der Neologismenstatus individueller Phraseme verworfen werden, es ist aber in der Regel nicht möglich, diesen Status definitiv zu bestätigen. Bessere Hinweise liefert eine Kombination der vorgestellten Methoden. Alle vier Methoden können zu einem gewissen Grad mit elektronischer Unterstützung ausgeführt werden, insbesondere der Abgleich des Phraseminventars vorhandener Beschreibungen und Korpora.

Im Folgenden wird ein solcher computergestützter Abgleich zweier idiomatischer Sammlungen zur Ermittlung potentieller phraseologischer Neologismen beschrieben. Zur Validierung werden vorhandene Ressourcen konsultiert sowie eine Informantenbefragung durchgeführt.

Die Idee eines solchen Abgleichs ist nicht neu, bereits Piotrowski (2009: 112) hat – allerdings für die Identifizierung von lexikalischen Neologismen – einen (halb-)automatischen Vergleich von Wörterbüchern oder Wortlisten vorgeschlagen. Im vorliegenden Beitrag werden der *Duden 11* und der *redensarten-index.de* verwendet, wobei davon ausgegangen wird, dass sich deren Inhalte aus diversen Gründen unterscheiden (siehe Abschnitt 3.1) und somit die Identifikation von phraseologischen Neologismen möglich ist.

Nachdem per computergestütztem Abgleich potentielle phraseologische Neologismen identifiziert wurden, gilt es, diese zu validieren. Dazu werden, wie bereits beschrieben, vorhandene idiomatische Sammlungen und Wörterbücher konsultiert sowie Textkorpora im Konsultationsparadigma ausgewertet. Wird durch einen dieser Schritte ein Beleg für ein Vorkommen eines potentiellen phraseologischen Neologismus vor dem Jahr 2000 gefunden, kommt dieser nicht weiter als Neologismus infrage. Abschließend wird eine Informantenbefragung durchgeführt (siehe Abschnitt 4.2), um den Bekanntheitsgrad der verbliebenen potentiellen Neologismen zu ermitteln und zu bestimmen, ob diese Einheiten von den Probanden als Neologismen eingeschätzt werden.

Diejenigen potentiellen phraseologischen Neologismen, die allen Validierungsschritten genügen, gelten auf der Grundlage unserer Methodik als (echte) phraseologische Neologismen.

Im Folgenden wird das Vorgehen ausführlich beschrieben, wobei auch die Schwierigkeiten aufgezeigt werden, welche die Identifikation und die Validierung von phraseologischen Neologismen mit sich bringen.

3 Ermittlung potentieller phraseologischer Neologismen

3.1 Material

Für die Ermittlung potentieller phraseologischer Neologismen vergleichen wir zunächst das Inventar einer umfangreichen aktuellen Online-Sammlung (*redensarten-index.de*) und einer aktuellen gedruckten Sammlung (*Duden 11*).

3.1.1 Online-Sammlung

Der *redensarten-index.de* ist eine idiomatische Sammlung, die seit 2001 vom Hobby-Phraseologen Peter Udem² betrieben wird. Gemäß Korrespondenz mit dem Betreiber vom 14.8.2011 werden Wendungen aufgenommen und erklärt, die für einen Fremdsprachler als erklärungsbedürftig betrachtet werden. Die Bearbeitung (d.h. die Entscheidung, welche Wendungen aufgenommen werden, und das Verfassen der Bedeutungsangaben sowie ggf. von Beispielsätzen) übernimmt ausschließlich der Betreiber. Es handelt es sich um eine Online-Sammlung, was theoretisch unbegrenzte Datenmengen sowie eine schnelle und flexible Erweiterung des Phrasenbestands zulässt. Zum Zeitpunkt unseres Abgleichs enthielt der *redensarten-index.de* 12.235 Einträge (Stand vom 16.8.2011).

Den fortlaufend erweiterten Bestand bezieht der *redensarten-index.de* aus vier Quellen: (1) Listen eines privaten Sammlers (die in den 1960ern begonnen wurden, jedoch nicht mehr fortgeführt werden), (2) Listen des Betreibers der Website, mit denen gemäß Korrespondenz (s.o.) die subjektiv empfundenen "einschlägigen Redewendungen aus dem Gebrauch der deutschen Sprache" gesammelt werden, (3) den Vorschlägen der Nutzer und (4) den Suchanfragen der Nutzer. Als Rechercheinstrument dient beim *redensarten-index.de* eine Internetrecherche mit Google, womit überprüft wird, ob die für eine Aufnahme infrage kommenden Wendungen "tatsächlich existieren und verwendet werden" (ebd.). Das "Nachschlagen in Wörterbüchern wie dem Duden oder Röhrich" wird "allenfalls als Ergänzung" zur Google-Abfrage vorgenommen (ebd.). Dabei reicht für die Aufnahme bereits ein relativ geringer Schwellenwert von rund 100 Treffern aus. Die von Juska-Bacher et al. (im Druck) und Kilgariff (2007) formulierten Vorbehalte gegenüber der Nutzung von Google zur Korpusarbeit müssen hier natürlich auch als generelle Kritik am Vorgehen angebracht werden.

3.1.2 Gedruckte Sammlung

Der *Duden 11* ist die gegenwartsnächste gedruckte idiomatische Sammlung (erschieden 2008), die nach wissenschaftlichen Kriterien erstellt wurde. Ziel der Sammlung ist, "die gebräuchlichsten festen Wendungen der deutschen Gegenwartssprache möglichst umfassend zu dokumentieren." (Duden 11 2008: 15). Auf dem Umschlag ist vermerkt, die Sammlung enthalte "[m]ehr als 10.000 feste Wendungen, Redensarten und Sprichwörter".

2 Wir danken dem Betreiber des *redensarten-index.de*, Peter Udem, für das schnelle und freundliche Beantworten aller Fragen und das Einverständnis, mit seinen Daten zu arbeiten. Unser Dank geht ebenso an die Dudenredaktion für die Bereitschaft, uns Auskünfte zu erteilen.

Gemäß Korrespondenz mit der Duden-Redaktion vom 15.8.2011 wird der *Duden 11* von einem zweiköpfigen Redaktionsteam erstellt, wobei abgesehen von den im Werk aufgeführten Quellen (*Duden 11* 2008: 15–16 und 921–958) die Duden-Sprachkartei und das aktuelle Dudenkorpus verwendet wurden. Die Duden-Sprachkartei wurde bis etwa 2005 geführt und enthält ca. zwei Millionen Karteikartenexzerpte aus dem 20. Jahrhundert mit Schwerpunkt auf der Nachkriegszeit; das aktuelle Dudenkorpus enthält ca. zwei Milliarden laufende Wortformen und setzt sich – wie auch die Sprachkartei – aus literarischen Texten, Zeitungs- und Zeitschriftentexten sowie Texten aus Sach- und Fachbüchern zusammen.

Die Quellen des *Duden 11* sind also auf gedruckte Texte konzentriert, mündliche Sprache ist entsprechend untervertreten. Jedoch geht man in der Phraseologie davon aus, „dass die Mündlichkeit dominiert, d.h. dass Phraseologismen in der gesprochenen Alltagssprache wesentlich häufiger auftreten als in geschriebenen Texten“ (Quasthoff et al. 2010: 37), wobei die Jugendsprache einen besonderen Stellenwert besitzt (Burger 2010: 131; Ehrhardt 2007: 259). Eine empirische Überprüfung dieser These steht bislang aus, denn „[l]eider gibt es bisher keine genügend großen Korpora gesprochener Sprache, um korpusbasierte Vergleiche der mündlichen und schriftlichen Phraseologie durchzuführen.“ (Quasthoff et al. 2010: 38). Wörterbücher haben die Schwäche, „nicht die aktuelle Sprachsituation [...], nicht den tatsächlichen Sprechakt“ (Röhrich 1991: 47) zu bezeugen.

3.1.3 Gegenüberstellung

Auch der *redensarten-index.de* bildet die mündliche Sprache nicht ab, er erlaubt es jedoch deutlich eher als der *Duden 11*, auch konzeptionell mündliche Sprache (zur Begrifflichkeit siehe Koch/Oesterreicher 1986) zu erfassen, weil die Überprüfung neuer Phraseme über eine Internetsuchmaschine erfolgt, die einen großen Anteil konzeptionell mündlicher Texte berücksichtigt (z.B. in Foren und Blogs). So kann der Nutzer auf zwei Ebenen Einfluss nehmen: Einerseits fließen die von ihm in Internettexten verwendeten oder in Foren erfragten Phraseme ein, andererseits kann er durch Vorschläge oder Anfragen an den *redensarten-index.de* eine Aufnahme in diesen direkt steuern.

Dem *Duden 11* steht diese Möglichkeit aufgrund der Veröffentlichung als gedruckte Sammlung nicht zur Verfügung.³ Ein weiterer Vorteil des *redensarten-index.de* für die Aufnahme von phraseologischen Neologismen besteht darin, dass er aufgrund seiner Konzeption als ausschließlich elektronisches Nachschlagewerk

³ Gemäß Korrespondenz mit der Dudenredaktion (ebenfalls vom 15.8.2011) fließen auch „Hinweise der Benutzer früherer Ausgaben“ ebenso ein wie „eigene Funde bei der täglichen lexikographischen Arbeit für andere Projekte“, doch bietet eine elektronische Suchmaske wie die des *redensarten-index.de* ganz andere Möglichkeiten bei der Suche nach Varianten oder Neologismen.

jederzeit überarbeitbar und erweiterbar ist und so laufend aktualisiert werden kann. Zudem erfolgt die Überprüfung von zur Aufnahme in den *redensarten-index.de* infrage kommenden Phrasemen durch Internetrecherche mit einer tiefen Treferschwelle, so dass Phraseme in ihrem Entstehungsprozess schon früh erfasst werden können. Und schließlich braucht sich eine Online-Sammlung wegen ihres theoretisch nahezu unbegrenzten Speicherplatzes nicht wie der *Duden 11* auf die „gebräuchlichsten Wendungen“ (s.o.) zu beschränken.

Die Arbeit am *Duden 11* war vor dem Veröffentlichungszeitpunkt 2008 abgeschlossen, während die Arbeit am *redensarten-index.de* kontinuierlich fortgesetzt wird. Somit ist ein Abgleich des Inventars beider Sammlungen geeignet, um phraseologische Neologismen zu identifizieren. Phraseme, die im *redensarten-index.de* enthalten sind, jedoch nicht im *Duden 11*, werden in dieser Studie als potentielle Neologismen betrachtet.

Selbstverständlich kann nicht ausgeschlossen werden, dass auch der *Duden 11* phraseologische Neologismen enthält. Diese sind durch ein analog zu unserem hier durchgeführten Verfahren zu ermitteln, indem das Inventar des *Duden 11* mit einer älteren Sammlung abgeglichen würde.

3.2 Computergestützter Abgleich

Für den computergestützten Abgleich wurde am 16.8.2011 eine Liste aller im *redensarten-index.de* verzeichneten distinkten Wendungen extrahiert; dies sind 12.235 Einträge. Für den *Duden 11* wurden 13.287 distinkte Einträge ermittelt; dies sind nicht nur die Stichwörter, sondern alle Phraseme, die an einer beliebigen Stelle angeführt werden. Ein direkter Vergleich der Phraseme ist nicht möglich, da beide Sammlungen verschiedenen Formulierungsgrundsätzen folgen.

Betrachten wir das Beispiel *Hand anlegen*. Im *redensarten-index.de* finden wir diese Einträge:

- (1) (a) *Hand an jemanden legen*
 (b) *Hand an sich legen*
- (2) (a) *letzte Hand an etwas legen*
 (b) *letzte Hand anlegen*

Im *Duden 11* folgende Einträge:

- (3) (a) *Hand an jmdn. legen*
 (b) *Hand an sich legen*
- (4) (a) *letzte Hand an etw. legen*
 (b) *letzte Hand anlegen*
- (5) *selbst Hand anlegen*

Obwohl für den Menschen auf den ersten Blick eindeutig ist, dass sich (1a) und (3a) entsprechen – sie unterscheiden sich lediglich in der Notation der zu füllenden Objektstelle – und (2a) und (2b) Varianten voneinander sind, ist dies für eine automatische Prüfung nicht erkennbar.

Im *Duden 11* finden wir:

(6) *sich den Teufel um etw. scheren*

Im *redensarten-index.de*:

(7) *sich den Teufel um etwas scheren / kümmern*

Auch diese beiden Einträge sind äquivalent. Neben der schon angesprochenen Unterschiedlichkeit bei der Angabe von Valenzstellen, finden wir hier das Phänomen der Variantenschreibung. Innerhalb eines Eintrags können Varianten für einzelne Elemente des Phrasems explizit angegeben sein, oder es kann für jede Variante einen gesonderten Eintrag geben. Die automatische Auflösung dieser Fälle ist ebenfalls nicht möglich.

Selbst innerhalb des *Duden 11* finden sich verschiedene Formulierungen. So treten sowohl *Geben ist seliger als/denn Nehmen* als auch *Geben ist seliger denn/als Nehmen* auf. Im Prozess der automatischen Deduplizierung werden solche Fälle nicht als äquivalent erkannt.

Der Abgleich dieser beiden Sammlungen basiert daher auf dem Vergleich der in den Einträgen enthaltenen Substantive und Verben. Mit *Stripey Zebra*,⁴ einem Werkzeug zur automatischen morphologischen Analyse, werden für jedes Phrasem alle Wortformen analysiert; jede Wortform wird auf ihre Grundform zurückgeführt, die Wortart und die morphologischen Eigenschaften werden ermittelt. Nur Verben und Substantive werden weiterbehandelt, alle anderen Wortformen nicht. Zur Qualität der Resultate automatischer morphologischer Analysesysteme für Deutsch sei auf Mahlow/Piotrowski (2009) verwiesen.

Mit einem Perl⁵-Programm werden die lemmatisierten Verben und Substantive als Schlüssel gespeichert, die zugehörigen Phraseme als Wert. Ist ein Schlüssel (also eine Substantiv- oder Verbgrundform) bereits vorhanden, wird das Phrasem zu dem/den bereits vorhandenen hinzugefügt. Das Resultat dieses Abgleichs ist eine alphabetische Liste für den *redensarten-index.de* und eine Liste für den *Duden 11* mit den erkannten Substantiven und Verben, für die jeweils notiert ist, in welchen Phrasemen sie vorkommen.

⁴ *Stripey Zebra* ist die aktuelle Version der Deutschen Malaga Morphologie (DMM) (Lorenz 1996) und liefert detaillierte strukturierte Analyseergebnisse (siehe auch Mahlow/Piotrowski 2009).

⁵ *Perl* ist eine Programmiersprache (siehe dazu <http://www.perl.org/>).

Ein Beispiel zur Veranschaulichung:

- (8) Für *jemandem Sand in die Augen streuen* werden “Sand”, “Augen” und “streuen” als relevante Grundformen ermittelt; das Phrasem wird für alle drei jeweils als Wert gespeichert.
- (9) Für *jemandem Salz in die Wunde streuen* (Nennform nach *redensarten-index.de*) werden “Salz”, “Wunde” und “streuen” ermittelt; dieses Phrasem wird anschließend zu dem bereits vorhandenen für “streuen” gespeichert und zudem als Wert für “Wunde” und “Salz”.

Anschließend kann – wiederum mit einem Perl-Programm – automatisch geprüft werden, welche Lemmata der einen Liste nicht in der anderen Liste enthalten sind, die entsprechenden Phraseme werden ausgegeben. Dies ergibt 2.066 Lemmata mit insgesamt 2.577 Wendungen, die nicht im *Duden 11*, dafür aber im *redensarten-index.de* enthalten sind. Dabei werden manche Wendungen mehrfach aufgeführt, wenn diese mehr als ein Substantiv oder Verb enthalten, die im *Duden 11* nicht vorkommen, und entsprechend unter mehreren Grundformen in der Liste abgespeichert werden.

3.3 Ergebnisse des Abgleichs

Diverse Einträge der aus dem Abgleich hervorgegangenen Wendungen aus dem Inventar des *redensarten-index.de* entsprechen nicht der hier verwendeten, Burger (2010: 15–32) folgenden Definition von Phrasemen als polylexikalisch, idiomatisch und fest. Die Festigkeit kann an dieser Stelle nicht überprüft werden (siehe dazu Abschnitt 4.2), allerdings kann davon ausgegangen werden, dass alle Phraseme über eine gewisse Festigkeit verfügen, da sie im Internet recherchiert wurden, bevor sie in den *redensarten-index.de* aufgenommen worden sind.

Ob die Wendungen den Kriterien Polylexikalität und Idiomatizität entsprechen, lässt sich relativ leicht feststellen. Monolexikalische Einträge wie *Ahoi!*, *Brummi* oder *Gedöns* wurden manuell aussortiert, weil diese das Kriterium der Polylexikalität nicht erfüllen. Dies traf auf ca. 1.500 Einträge zu, darunter ca. 800 Komposita wie *Angstphase*, *Dreckspatz* oder *Hupkonzert*. Ebenfalls aus der Liste gestrichen wurden nicht oder nur schwach idiomatische Wendungen (z.B. *keinen Benimm haben*).

Die weitere manuelle Bereinigung verfolgt das Ziel, alle Phraseme auszuschließen, die aufgrund der uneinheitlichen Ausrichtungen, Erstellungs- und

Formulierungsmethoden (siehe Abschnitt 3.1) von *redensarten-index.de* und *Duden 11* unerwünscht den Abgleich überstanden haben. Dies gilt für Phraseme, die vom *redensarten-index.de* und vom *Duden 11* in uneinheitlicher Nennform aufgeführt werden (siehe Beispiele (1) bis (6)), sowie für Varianten, die sich bloß durch Substitution semantisch ähnlicher Elemente unterscheiden, sofern nicht sowohl im *Duden 11* als auch im *redensarten-index.de* die gleiche Variante verzeichnet ist.

So kennt der *redensarten-index.de* beispielsweise nur die Form *sich kringeln vor Lachen/Vergnügen* ("sich köstlich amüsieren, heftig lachen"), der *Duden 11* hingegen die Formen *platzen/sich biegen vor Lachen* ("unbändig heftig lachen") sowie *sich kring(e)lig lachen* (ohne Bedeutungsangabe). Wegen der lexikalischen Ähnlichkeit in Verbindung mit den Bedeutungsangaben werden diese Phraseme als Varianten betrachtet. Damit kommt *sich kringeln vor Lachen/Vergnügen* als phraselogischer Neologismus nicht mehr in Frage, was automatisch nicht feststellbar ist.

Ein systematisches Verfahren zum Ausschluss von Varianten ist komplex, denn dies erfordert gerade bei Neologismen zu große Kompetenzen des Bearbeiters, weil nicht nur die Kenntnis des Phrasems, sondern auch die Kenntnis der Bedeutungsabstufungen weder vorausgesetzt noch ohne weiteres erschlossen oder nachgeschlagen werden kann. Zudem sind die Grenzen zwischen Variante und eigenständigem Phrasem oftmals fließend. Einzig bei Varianten, die durch paradigmatische Substitution einzelner Lexeme entstanden sind, scheint der Ausschluss eines Phrasems aus der Gruppe der potentiellen Neologismen methodisch vertretbar. So kennt z.B. der *Duden 11* das Phrasem *wissen, wo Barthel den Most holt* ("alle Kniffe kennen") und der *redensarten-index.de* die Variante *wissen/zeigen, wo der Opa/Bauer den Most holt*. ("gewitzt/schlau sein; [. . .]").

Zudem muss berücksichtigt werden, dass eine automatische Lemmatisierung auch auf Elemente des Phrasems wirkt, die nicht zu den obligatorischen Elementen zählen. Ein Beispiel hierfür ist der Eintrag *böses Blut machen/schaffen* des *Duden 11*. Im *redensarten-index.de* ist das Phrasem in der Form *böses Blut erzeugen* verzeichnet. Den festen Kern des Phrasems bildet *böses Blut*, das Verb lässt offensichtlich eine gewisse Variabilität zu. In solchen Fällen steigt die Fehleranfälligkeit des automatischen Abgleichs durch Lemmatisierung: zwar werden beide Phraseme dem Lemma *Blut* zugeordnet, das *Duden*-Phrasem jedoch zusätzlich den Lemmata *schaffen* und *machen*, das *redensarten-index*-Phrasem dem Lemma *erzeugen*.

Von den durch den Abgleich vom *redensarten-index.de* und dem *Duden 11* ermittelten Wendungen verbleiben nach der manuellen Bereinigung noch 1.006 Phraseme als potentielle Neologismen.

4 Validierung potentieller phraseologischer Neologismen

In der Validierung werden die im Identifikationsverfahren (siehe Abschnitt 3) ermittelten potentiellen phraseologischen Neologismen aussortiert, für die in verschiedenen idiomatischen Sammlungen, Wörterbüchern und Korpora Belege aus der Zeit vor dem Jahr 2000 gefunden werden und welche somit gemäß unserer Definition keine Neologismen sind. Anschließend wird durch eine Informantenbefragung versucht, echte Neologismen von solchen Pseudoneologismen zu unterscheiden.

4.1 Auswertung von Textkorpora, idiomatischen Sammlungen und Wörterbüchern

Als Korpora bzw. Textsammlung wurden die online verfügbaren Textkorpora *C4* (mit Texten aus dem 20. Jahrhundert), *Cosmas II* (mit Texten bis zur Gegenwart) und – um auch ein mündliches Korpus zu konsultieren – das *Deutsche Spracharchiv* (DSAv) sowie der Band 125 der Digitalen Bibliothek *Von Luther bis Tucholsky* verwendet.

Zu den konsultierten idiomatischen Sammlungen gehören das *Deutsche Sprichwörter-Lexikon* von Karl Friedrich Wilhelm Wander (2001) sowie das *Lexikon der sprichwörtlichen Redensarten* von Lutz Röhrich (2004). Ebenfalls hinzugezogen wurden das *Variantenwörterbuch des Deutschen* (Ammon et al. 2004), das Neologismenwörterbuch *Neuer Wortschatz* (Herberg et al. 2004) und – erneut um die mündliche Sprache stärker zu berücksichtigen – das *Wörterbuch der deutschen Umgangssprache* (Küpper 2004).

Bei dieser Prüfung zur Validierung der Phraseme als Neologismen war entscheidend, ob – und falls ja – ab wann die Phraseme belegt sind. Wurden Belege von vor 2000 gefunden, handelt es sich bei den entsprechenden Phrasemen gemäß unserer Definition (siehe Abschnitt 1) nicht um Neologismen. Für die meisten der potentiellen Neologismen wurden solche Belege gefunden.

Diejenigen 79 Phraseme, die das Verfahren bis hierher überdauert haben, werden in einem weiteren Schritt mit Hilfe eines Fragebogens validiert.

4.2 Informantenbefragung

Um die verbliebenen potentiellen phraseologischen Neologismen auf ihre überregionale Gebräuchlichkeit und ihren Neuheitswert zu prüfen, wurden im

Dezember 2011 MuttersprachlerInnen aus unterschiedlichen Regionen des deutschen Sprachraums zu ihrer Kenntnis und Alterseinschätzung der 79 Phraseme befragt (Phrasemliste siehe Anhang 1). An dieser Befragung beteiligten sich 41 Personen aus Deutschland, Österreich und der Schweiz. Für das Testen dieser Methode zur Validierung potentieller Neologismen scheint uns dieses kleine Sample ausreichend, eine genauere Beschreibung der regionalen Verteilung der einzelnen Phraseme lässt sich daraus selbstverständlich nicht ableiten.

Die verschiedenen Bundesländer bzw. Kantone wurden zu Großräumen zusammengefasst (zur räumlichen Verteilung der Informanten siehe Anhang 2). Aus jeder Region wurde mindestens eine Person pro Altersgruppe (Altersstufe I: zwischen 20 und 30 Jahren, Altersstufe II: zwischen 50 und 60 Jahren) befragt, um nicht nur eventuelle regionale Unterschiede ausmachen zu können, sondern auch überprüfen zu können, ob phraselogische Neologismen jungen Personen eher bekannt sind und damit aus der Jugendsprache stammen könnten.

Beide Personengruppen füllten denselben Fragebogen aus. Dieser enthielt 79 Phraseme (Ergebnis der bisherigen Identifikations- und Validierungsschritte) mit Bedeutungsangaben (siehe Anhang 1).

Für jedes Phrasem in dieser Liste standen den Informanten drei Antwortmöglichkeiten zur Verfügung: Sie konnten ankreuzen, (1) dass ihnen das Phrasem unbekannt war, (2) dass das Phrasem ihrer Meinung nach bereits vor 2000 bekannt war oder (3) dass das Phrasem ihrer Meinung nach erst ab 2000 entstanden ist.

Zusätzlich hatten die Informanten die Möglichkeit, ihnen bekannte und von der Liste abweichende Phrasemformen zu ergänzen oder anderweitige Bemerkungen anzubringen. Notiert wurden:

- vermeintliche regionale Begrenzungen von Phrasemen (eine Informantin aus Schleswig-Holstein notierte zu *Wat mutt, dat mutt!* “norddeutsch”),
- Angaben zum Alter von Phrasemen (ein Informant aus Niedersachsen bemerkte zum Phrasem *jemandem ein Ei auf die Schiene nageln*, er glaube, es etwa im Jahr 1990 im Film *Didi auf vollen Touren* (1986, Regie: Wigbert Wicker) gehört zu haben),
- andere Bedeutungen eines Phrasems (einem Informanten ist das Phrasem *auf einen Trompetenkäfer treten* weniger in der Bedeutung “einen Darmwind ablassen”, sondern “eher als Ausdruck für einen Rülpsen” bekannt),
- Varianten von Phrasemen (*um die Ecke lunschen* statt [. . .] *luchsen*),
- und synonyme Phraseme (*mit der Kneifzange gepudert sein* für *vom Lemmes gepickt sein*).

Diese Möglichkeit, Kommentare zu Phrasemen anzubringen, wurde nicht häufig genutzt, doch zeigt sich hier generell, welche Art Informationen durch eine

Potentielle phraseologische Neologismen	Bedeutung nach <i>redensarten-index.de</i>
der fleischliche Akt	der Geschlechtsakt/Koitus
Man soll die Äpfel erst zählen, bevor man sie aufteilt.	Man soll nicht voreilig handeln/erst Fakten sammeln.
einen tollen Balkon haben	einen schönen Busen haben
Gleich klatscht es, aber kein* Beifall! [*Nennform nach <i>redensarten-index.de</i>]	Androhung von Gewalt
einen Bob in die Bahn jagen/schicken/ legen/werfen/ setzen	Stuhlgang haben; koten; defäkieren
einen Clown zum Frühstück gegessen/ verspeist haben	unangebracht lustig/aufdringlich witzig sein
Dann ist der Drops gelutscht!	Dann ist die Sache (erfolgreich oder erfolglos) erledigt/fertig/abgeschlossen.
um die Ecke lutschen	heimlich/verstoßen gucken
jemandem ein Ei auf die Schiene nageln	jemandem etwas Schwieriges vorsetzen; jemanden nasführen/anführen; jemandem Probleme bereiten
sich ein Ei drauf braten/pellen/backen	sich von Vorwürfen nicht stören lassen; Kritik ignorieren; gleichgültig sein
Am Ende/hinten kackt die Ente.	Am Ende kommt die große Entscheidung; der Sieger steht erst am Ende fest; man sollte sich nicht zu früh freuen.
einen Esel zu kämmen haben	Verpflichtungen/zu tun haben
bis ins Essgefach	bis ins Detail; bis in den letzten Winkel der Privatsphäre; überall hin
Froschblut in den Adern haben	gefühlskalt/lethargisch sein
gutes Geld schlechtem hinterherwerfen	trotz einer Fehlinvestition weiteres Geld nachschießen; noch einmal Geld ausgeben müssen, um Forderungen einzutreiben
eine alte Gewitterhexe	zänkische, böse Frau
die Güte des Geschicks	die glückliche Wendung der Ereignisse; günstige Umstände
die Hacken in den Teer hauen	sich beeilen; sich anstrengen
Egal ist ein Handkäs, der stinkt von allen/ beiden Seiten!	Abwehr von "Das ist mir egal!"
jemanden durch die Hecke ziehen	koitieren; Geschlechtsverkehr haben; mit einer Frau schlafen
Fick die Henne!	Fluch; ruft man aus, wenn etwas richtig schief geht
fast einen Herzkasper kriegen	sich heftig erschrecken
nicht ganz richtig im Hinterstübchen sein	leicht verrückt sein
Jemandem haben sie das Hirn/Gehirn amputiert.	jemand ist sehr dumm/verrückt.
sich/jemanden zum Horst machen	sich/jemanden blamieren/lächerlich machen
kurze Hose, Holzgewehr	sich wie ein kleiner Junge fühlen: unvorbereitet, schlecht informiert, hilflos

Potentielle phraselogische Neologismen	Bedeutung nach <i>redensarten-index.de</i>
die Hühner satteln	aufbrechen; sich zum Gehen fertig machen
wie Hulle	sehr; sehr stark
einen nassen Hut aufhaben	leicht verrückt/nicht recht bei Verstand sein; spinnen
über die Hutkrempe gehen	übersteigen; überschreiten; darüber hinausgehen
Es gibt Kasalla.	Es gibt Ärger/Prügel/Krawall.
noch als Käse im Schaufenster gelegen haben	zu dieser Zeit noch nicht geboren worden sein
den Kasper schnäuzen	onanieren; masturbieren
Privat geht vor Katastrophe.	Die Freizeit ist wichtiger, auch wenn der Betrieb in einer Notlage ist.
rubbel die Katz	schnell; ohne Umschweife; schwungvoll
jemandem die Kauleiste tiefer legen	jemanden schlagen/verprügeln
auf Kies gefurzt haben	sich bei etwas verrechnet/verschätzt haben; mit etwas nicht gerechnet haben; etwas nicht bedacht haben
dumm sein wie ein Konsumbrot	sehr dumm sein
jemandem ein Kotelett an die Backe labern/ schwätzen/ quatschen/sabbeln	jemanden beschwatzen; viel reden; jemandem mit seinem Gerede auf die Nerven gehen
Mir kommt die (kalte) Kotze hoch.	sagt man, um Unverständnis/Ablehnung/ Missbilligung/Überdruß auszudrücken
Wer kriecht, stolpert nicht.	Wer sich anpasst, dem wird nichts Schlimmes passieren; ein selbstbestimmtes Leben ist riskant.
Krokodilstränen vergießen	Trauer/Schmerz heucheln
den Kümmel aus dem Käse pulen/ziehen/ krümeln	etwas bis ins Kleinste auseinander nehmen/ analysieren; penibel/kleinlich/pedantisch/ übermäßig kritisch sein; nörgeln
sich vor Lachen wegwerfen	heftig lachen
viel Lametta auf der Brust haben	viel Orden/Ehrenzeichen besitzen
wie Latsch und Bommel	zusammengehören; untrennbar sein; gut zusammenpassen
vom Lemmes gepickt sein	leicht verrückt/im Kopf nicht ganz richtig sein
den Lurch würgen	onanieren; masturbieren
Wat mutt, dat mutt!	Was sein muss, muss sein; es ist notwendig; was getan werden muss, das muss getan werden.
(Na dann) prost Nägeli!	Sagt man, wenn man nichts Gutes erwartet.
mit etwas eine Notlandung machen	um totalen Misserfolg knapp herumkommen
Der Ober sticht den Unter.	Derjenige in der höheren Position setzt sich durch; er herrschen hierarchische Verhältnisse.
jemanden/etwas vor die Pumpe laufen lassen	jemanden ignorieren/auflaufen lassen; die Unterstützung verweigern; etwas zum Scheitern bringen
Riemen- und Dollenbruch!	einem Ruderer beim Wettkampf viel Glück und Erfolg wünschen
Sabbelwasser getrunken haben	viel reden

Potentielle phraseologische Neologismen	Bedeutung nach <i>redensarten-index.de</i>
Grün und blau schmückt die Sau.	Grün und blau passen nicht zusammen; grün-blau ist keine harmonische Farbkombination.
Geh scheißen/kacken!	Verschwinde! Geh weg!
nicht viel Schlamm in der Kuhle haben	nicht intelligent sein
Geh gleich zum Schmied und nicht zum Schmiedl; geh nicht zum Schmiedl, sondern gleich zum Schmied; man geht besser zum Schmied als zum Schmiedl.	Man solle sich gleich an den Richtigen/den Experten/den Spezialisten wenden.
Etwas ist Schmuck am Nachthemd.	Etwas ist überflüssig/sinnlos.
sich freuen wie ein Schnitzel	sich sehr freuen
Ein Schuft, der Böses dabei denkt seinen Sermon (he)runterbegen/ (he)runterleiern	Ein Schurke, der Böses dabei denkt einen langweiligen Vortrag halten; sich ständig wiederholen
Jemandem scheint die Sonne aus dem Arsch jemandem geht der Stift.	Jemandem geht es gut; jemand hat gute Laune Jemand hat Angst.
auf Trallafitti gehen	ausgehen; feiern gehen
auf einen Trompetenkäfer treten	Blähungen haben; einen Darmwind abgehen lassen; pupsen
durch den/in Tüdel kommen	etwas durcheinander bringen; sich irren
Er/sie (/du) kann(st) mich mal am Tüffel tuten!	Ausruf, wenn man mit etwas/jemandem nichts zu tun haben möchte; Ausdruck von Überdruß/Ablehnung
Du bist wohl in der U-Bahn geboren!	Sagt man, wenn jemand nicht die Türe hinter sich schließt.
Holla die Waldfee!	Ausruf des Erstaunens/der Anerkennung (bei eventueller gleichzeitiger subtiler Andeutung, dass der Sachverhalt nicht sehr wichtig oder sogar lächerlich ist)
fünf gegen Willi	Onanie; Masturbation
Leck die Ziege!	Ausdruck von Überraschung/Verwunderung/ Bekräftigung
auf Zinne sein	wütend sein
eine olle Zippe	Schimpfwort
Der Zitronenfalter faltet ja auch keine Zitronen!	Sagt man, um auszudrücken, dass man einen Begriff nicht zu wörtlich nehmen sollte; sagt man, wenn jemand seiner Funktion nicht gerecht wird.
den Zonk ziehen	eine schlechte Wahl treffen; eine Niete ziehen; Pech haben
Der Zug hält an jeder Milchkanne.	Nahverkehrszug, der auch an kleinen Bahnhöfen hält
auf den Zwutsch/Swutsch gehen	ausgehen; feiern gehen; einen draufmachen; einen Kneipenbummel machen; eine Diskothek besuchen

Land	Region	Bundesländer bzw. Kantone
Deutschland	Nordwesten	Bremen, Hamburg, Niedersachsen, Schleswig-Holstein
	Nordosten	Berlin, Brandenburg, Mecklenburg-Vorpommern, Sachsen-Anhalt
	Mitte, Westen	Hessen, Nordrhein-Westfalen
	Mitte, Osten	Sachsen, Thüringen
	Südwesten	Baden-Württemberg, Rheinland-Pfalz, Saarland
	Südosten	Bayern
Schweiz	Nordwesten	Aargau, Baselland, Baselstadt, Solothurn
	Nordosten	Appenzell Ausserrhoden, Appenzell Innerrhoden, Schaffhausen, St. Gallen, Thurgau, Zug, Zürich
	Südwesten	Bern, Freiburg, Luzern
	Südosten	Glarus, Graubünden, Nidwalden, Obwalden, Schwyz, Uri
Österreich	Osten	Burgenland, Niederösterreich, Wien
	Westen	Rest Österreichs

Anhang 2: Die regionale Verteilung der Informanten

Befragung gesammelt werden können. Für phraseologische Neologismen sind besonders die Einschätzung des Alters von Interesse wie auch Bedeutungsangaben (durch Abweichungen zu den auf dem Bogen notierten Angaben oder synonyme Phraseme). Theoretisch können die von den Informanten angegebenen Varianten und Synonyme in Belegsammlungen (siehe Abschnitt 2.1) einfließen, die es dann wiederum zu verifizieren gilt.

Am Ende des Fragebogens wurden die Informanten gebeten, weitere vermeintliche phraseologische Neologismen zu notieren, die ebenfalls den Validierungsschritten unterzogen werden könnten. Diese Möglichkeit wurde von den Informanten jedoch nur vereinzelt genutzt. Zudem bestand ein nicht unerheblicher Teil der hier genannten Phraseme aus bereits seit längerer Zeit etablierten Einheiten (z.B. *umgekehrt wird ein Schuh daraus*). Diese Art der Beteiligung von Informanten an der Identifizierung von Neologismen erwies sich in unserer Studie als weniger geeignet.

4.3 Ergebnisse

Die Bewertung der einzelnen potentiellen phraseologischen Neologismen durch die Informanten fiel sehr unterschiedlich aus. Ein Phrasem (*Krokodilstränen vergießen*) war allen Informanten bereits vor dem Jahr 2000 bekannt, eines (*vom Lemmes gepickt sein*) war gänzlich unbekannt, die übrigen wurden gemischt eingestuft. Prinzipiell wurde die Einschätzung, dass ein Phrasem erst seit dem

Jahr 2000 bekannt und damit nach unserer Vorgabe ein Neologismus ist, selten getroffen (in 7% der Fälle, d.h. gut 220 Mal).

Die Klassifizierung der Neologismuskandidaten als (echter) Neologismus war an zwei Bedingungen geknüpft:

1. Das Phrasem musste eine Mindestbekanntheit von 20% aufweisen (Nachweis der Festigkeit in der Sprachgemeinschaft, Ausschluss von Regionalismen). Dieses Kriterium erfüllten 56 der insgesamt 79 Phrasemkandidaten.
2. In Anbetracht der schwierigen Datierbarkeit des Alters von Phrasemen genügte es, wenn mindestens ein Drittel der Informanten, die das Phrasem kannten, es als Neologismus einstufen, also als nach 2000 entstanden. Dieses Kriterium erfüllten noch 15 der 56 verbliebenen Phraseme, die wir daher als phraseologische Neologismen klassifizieren (siehe Tab. 1).

phraseologische Neologismen	Anzahl		
	unbekannt	vor 2000	nach 2000
einen Clown zum Frühstück gegessen/ verspeist haben	18	10	11
rubbel die Katz	18	11	11
sich/jemanden zum Horst machen	17	12	10
jemandem ein Kotelett an die Backe labern/ schwätzen/quatschen/sabbeln	21	10	9
Geh scheißen/kacken!	18	12	9
sich freuen wie ein Schnitzel	24	6	9
Dann ist der Drops gelutscht!	21	11	8
Fick die Henne!	29	2	8
jemandem die Kauleiste tiefer legen	25	7	8
Gleich klatscht es, aber kein Beifall!	22	11	7
Jemandem scheint die Sonne aus dem Arsch.	24	9	7
den Zonk ziehen	25	8	7
wie Hulle	29	5	5
Du bist wohl in der U-Bahn geboren!	25	10	5
jemandem geht der Stift.	27	8	4

Tabelle 1: Die mit diesem Verfahren ermittelten phraseologischen Neologismen (in abnehmender Folge der Häufigkeit der Einschätzung als Neologismus). Zu den Bedeutungen vgl. Anhang 1.

Trotz der verschiedenen Validierungsschritte ist keineswegs gesichert, dass es sich bei allen diesen Phrasemen um Neologismen handelt. Ein spezieller Fall liegt bei dem Phrasem *rubbel die Katz* vor, das gemäß der nach Zeitpunkt der

Aufnahme geordneten Redensartenliste bereits im Jahr 2004 in den *redensarten-index.de* aufgenommen wurde. Dieses Phrasem stellt auch den Titel eines Films von Detlef Buck dar, der am 15.12.2011 (kurz vor Beginn unserer Befragung) in deutschen Kinos anlief. Es ist möglich, dass einigen Informanten das Phrasem deshalb geläufig war und dadurch auch die Einschätzung des Alters beeinflusst wurde.

Frühere Untersuchungen zum Einfluss des Alters der Informanten auf ihre Phrasemkenntnis haben ergeben, dass ältere Informanten mehr der abgefragten Phraseme kennen als jüngere (z.B. Häcki Buhofer/Burger 1994: 15f. und Juska-Bacher 2006: 172f.). Allerdings handelte es sich in den betreffenden Untersuchungen bei den abgefragten Phrasemen um tradiertes phraseologisches Material. Wenn Neologismen – wie etwa Röhrich (1991: 40) oder Burger (2010: 131) annehmen – nicht selten aus der Jugendsprache stammen, wäre zu erwarten, dass zwischen den beiden befragten Altersgruppen deutliche Unterschiede zu erkennen sind. Informanten im Alter von 20 bis 30 Jahren sollten demnach eine größere Zahl der potentiellen Neologismen kennen als die Informanten im Alter von 50 bis 60 Jahren und sie entsprechend auch besser als Neologismen einzuschätzen vermögen.

Auf die Gesamtheit der 79 potentiellen phraseologischen Neologismen (siehe Anhang 1) bezogen trifft diese Erwartung nicht zu: Die Altersgruppe von 20 bis 30 Jahren gibt in 62% der Fälle an, ein Phrasem nicht zu kennen, während es bei der Gruppe der 50- bis 60-Jährigen nur geringfügig höhere 63% sind. Berücksichtigt man allerdings nur die 15 Phraseme, die nach unserem Verfahren als Neologismen eingestuft wurden (siehe Tabelle 2), ist der Unterschied deutlich größer: Der Anteil der als unbekannt angegebenen Phraseme reduziert sich bei der jüngeren Altersgruppe (20–30 Jahre) auf 46%, während er sich bei der Altersgruppe von 50 bis 60 Jahren auf 66% erhöht. Ein t-Test, der die Mittelwerte dieser Altersgruppen vergleicht, ergibt einen signifikanten Unterschied zwischen den Gruppen ($p = 0.04$).

Für die These, dass Neologismen nicht selten aus der Jugendsprache stammen, zeichnet sich anhand dieses sowohl hinsichtlich der Phraseme wie der Gewährspersonen kleinen Samples also trotzdem eine gewisse Tendenz ab: Die jüngere Probandengruppe kennt deutlich mehr Neologismen als die ältere Gruppe (54% gegenüber 34%).

5 Zusammenfassung

Phraseologische Neologismen sind schwer greifbar. Dies liegt einerseits an ihrer polylexikalischen Form, die mehr Variation zulässt, als es bei monolexikalischen

Neologismen oder auch etablierten Idiomen der Fall ist. Andererseits können sie systematisch nur in der schriftlich bezeugten Sprache nachgewiesen werden. Die mündliche Sprache ist wegen des Mangels an ausreichend aktuellem und umfangreichem Material nur schlaglichtartig zu erfassen, obwohl davon ausgegangen werden kann, dass phraseologische Neologismen oft zuerst in der gesprochenen Sprache auftreten.

Dies erschwert sowohl die Identifikation als auch die Validierung. Einen möglichen Lösungsansatz stellt die Kombination bewährter Methoden der Lexikographie wie Auswertungen vorhandener Beschreibungen, Informantenbefragungen, Belegsammlungen und Korpusanalysen dar. In diesem Artikel haben wir die Umsetzung beispielhaft gezeigt, um phraseologische Neologismen im Deutschen seit dem 1.1.2000 zu ermitteln. Dabei stützen wir uns auf eine eigene Definition phraseologischer Neologismen, die sich an herkömmliche Neologismusdefinitionen anlehnt.

Die Grundlage unseres Vorgehens bildet ein Abgleich des Inventars einer aktuellen gedruckten Sammlung, dem *Duden 11*, und einer laufend aktualisierten Online-Sammlung, dem *redensarten-index.de*. Der *Duden 11* wurde nach wissenschaftlichen Kriterien erstellt und kann als Standardwerk der synchronen Phraseographie angesehen werden; der *redensarten-index.de* ist eine kontinuierlich wachsende Sammlung, die nicht streng nach wissenschaftlicher Methodik vorgeht, jedoch Neologismen unbeschränkt und schneller aufnehmen kann als der *Duden 11*, und auch konzeptionell mündliche Sprache, etwa aus Foren oder Blogs, enthält.

Da ein direkter automatischer Abgleich des Inventars zweier Sammlungen aufgrund unterschiedlicher Erstellungs- und Formulierungsgrundsätze nicht möglich ist, verwenden wir computerlinguistische Ressourcen, um die Lemmata der enthaltenen Phraseme zu ermitteln und den Abgleich so über den Umweg des Abgleichs der beteiligten Elemente durchzuführen. Die Ergebnisse müssen anschließend manuell auf Eindeutigkeit und auf ihren Phrasemstatus überprüft werden.

Potentielle Neologismen werden in mehreren Schritten durch eine Kombination lexikographischer Methoden validiert: Die Ergebnisse des Abgleichs werden in Wörterbüchern, idiomatischen Sammlungen und Korpora recherchiert und hier belegte Phraseme ausgeschlossen. Anschließend werden die verbliebenen Einheiten durch eine Informantenbefragung auf ihre Bekanntheit und ihr vermeintliches Alter überprüft. Hier zeigt sich, dass die Einschätzung des Alters von Phrasemen große Unterschiede aufweist, jedoch die Methode der Informantenbefragung bei der für Neologismen so schwierigen Bedeutungsbestimmung wertvolle Hinweise liefert.

Durch diese Methodenkombination, die deutlich über bisher bei der Wörterbucharstellung verwendete Verfahren hinausgeht (z.B. Herberg et al. 2004)

konnten fünfzehn Phraseme der ursprünglich 2.577 automatisch ermittelten Kandidaten als phraselogische Neologismen gemäß unserer Definition bestimmt werden. Dazu konnten durch die Informantenbefragung weitere gesammelt werden, die es noch zu validieren gilt.

Es hat sich gezeigt, dass durch eine Kombination von etablierten lexikographischen Methoden und computergestütztem Vorgehen phraselogische Neologismen sowohl aufgespürt als auch validiert werden können. Aufgrund des Variantenreichtums der Phraseme ist eine automatische Ermittlung von Neologismen nicht möglich, eine manuelle Nachbearbeitung der Ergebnisse ist notwendig. Computerlinguistische Ressourcen können jedoch die Menge der manuell zu bearbeitenden Einheiten stark reduzieren und bringen damit einen erheblichen Vorteil.

Der Vergleich einer älteren und einer jüngeren Informantengruppe ergab, dass die jüngeren Personen mehr Neologismen kannten als ältere, was die Annahme stützt, dass Neologismen – wovon bspw. Burger (2010: 131) und Ehrhardt (2007: 259) ausgehen – zu einem nicht unerheblichen Teil aus der Jugendsprache stammen. Da letztere aber zu einem großen Teil zum Bereich konzeptionell mündlicher Sprache gehört, wiegt es besonders schwer, dass diese nicht adäquat untersucht werden kann.

Universität Basel, Switzerland

Bibliographie

Idiomatische Sammlungen & Wörterbücher

Ammon, Ulrich, Hans Bickel, Jakob Ebner, Ruth Esterhammer, Markus Gasser, Lorenz Hofer, Birte Kellermeier-Rehbein, Heinrich Löffler, Doris Mangott, Hans Moser, Robert Schläpfer, Michael Schloßmacher, Regula Schmidlin & Günter Vallaster. 2004. *Variantenwörterbuch des Deutschen: Die Standardsprache in Österreich, der Schweiz und Deutschland sowie in Liechtenstein, Luxemburg, Ostbelgien und Südtirol*. Berlin: Walter de Gruyter.

Dudenredaktion (Hrsg.). 2008. *Redewendungen*. Bearb. von Werner Scholze-Stubenrecht. Mannheim: Dudenverlag. Bd. 11.

Herberg, Dieter, Michael Kinne & Doris Steffens (Hrsg.). 2004. *Neuer Wortschatz. Neologismen der 90er Jahre im Deutschen*. Berlin: Walter de Gruyter.

Küpper, Heinz. 2004. *Wörterbuch der deutschen Umgangssprache* (CD-Rom, Digibib 36). Berlin: Directmedia Publishing.

Redensarten-Index.de. Online-Wörterbuch für Redensarten, Redewendungen, idiomatische Ausdrücke, feste Wortverbindungen. <http://www.redensarten-index.de/suche.php> Letzter Zugriff: 3.3.2012.

Röhrich, Lutz. 2000. *Lexikon der sprichwörtlichen Redensarten* (CD-Rom, Digibib 42). Berlin: Directmedia Publishing.

Wander, Karl Friedrich Wilhelm. 2001. *Deutsches Sprichwörter-Lexikon* (CD-Rom, Digibib 62). Berlin: Directmedia Publishing.

Sprachkorpora

COSMAS II. Corpus Search, Management and Analysis System. Online-Version. Cosmas2. ids-mannheim.de/cosmas2-web/Letzter Zugriff: 1.10.2011.

Deutsche Literatur von Luther bis Tucholsky. 2005. Digitale Großbibliothek (CD-Rom, Digibib 125). Berlin: Directmedia Publishing.

DSAv. Deutsches Spracharchiv und Datenbank Gesprochenes Deutsch. <http://dsav-oeff.ids-mannheim.de/> Letzter Zugriff: 1.10.2011.

Korpus C4. <http://www.korpus-c4.org/> Letzter Zugriff: 1.10.2011.

Forschungsliteratur

Bergenholtz, Henning & Joachim Mugdan. 1991. Formen und Probleme der Datenerhebung II: Gegenwartsbezogene synchronische Wörterbücher. In Franz Josef Hausmann, Oskar Reichmann, Herbert Ernst Wiegand & Ladislav Zgusta (Hrsg.), *Ein internationales Handbuch zur Lexikographie*, 1611–1625. Berlin/New York: Walter de Gruyter.

Burger, Harald. 2010. *Phraseologie: Eine Einführung am Beispiel des Deutschen*. Berlin: Erich Schmid Verlag.

Colson, Jean-Pierre. 2007. The World Wide Web as a corpus for set phrases. In Harald Burger, Dmitrij Dobrovol'skij, Peter Kühn & Neal R. Norrick (Hrsg.), *Phraseologie. Phraseology*, 1071–1077. Berlin/New York: Walter de Gruyter.

Ehrhardt, Claus. 2007. Phraseme in der Jugendsprache. In Harald Burger, Dmitrij Dobrovol'skij, Peter Kühn & Neal R. Norrick (Hrsg.), *Phraseologie. Phraseology*, 253–264. Berlin/New York: Walter de Gruyter.

Häcki Buhofer, Annelies & Harald Burger. 1994. Phraseologismen im Urteil von Sprecherinnen und Sprechern. In Barbara Sandig (Hrsg.): *EUROPHRAS 92, Tendenzen der Phraseologieforschung*, 1–33. Bochum: Brockmeyer.

Hallsteinsdóttir, Erla, Monika Šajánková & Uwe Quasthoff. 2006. Phraseologisches Optimum für Deutsch als Fremdsprache: Ein Vorschlag auf der Basis von Frequenz- und Geläufigkeitsuntersuchungen. In *Linguistik online* 27. 117–136.

Herberg, Dieter & Michael Kinne. 1998. *Neologismen*. Heidelberg: Gross.

Juska-Bacher, Britta. 2006. *Empirisch kontrastive Phraseologie: Am Beispiel der Bekanntheit der Niederländischen Sprichwörter im Niederländischen, Deutschen und Schwedischen*. Baltmannsweiler: Schneider Verlag Hohengehren.

Juska-Bacher, Britta, Chris Biemann & Uwe Quasthoff. Im Druck. Webbasierte linguistische Forschung: Möglichkeiten und Begrenzungen beim Umgang mit Massendaten. *Linguistik online*.

Kilgariff, Adam. 2008. Googleology is Bad Science. In *Computational Linguistics* 33(1). 147–151.

Koch, Peter & Wulf Oesterreicher. 1986. Sprache der Nähe – Sprache der Distanz: Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. In Olaf Deutschmann, Hans Flasche, Bernhard König, Margot Kruse, Walter Pabst & Wolf-Dieter Stempel (Hrsg.), *Romanistisches Jahrbuch*. Band 36, 15–43. Berlin/New York: Walter de Gruyter.

Lorenz, Oliver. 1996. *Automatische Wortformererkennung für das Deutsche im Rahmen von MALAGA*. Magisterarbeit an der Friedrich-Alexander-Universität Erlangen-Nürnberg.

Mahlow, Cerstin & Michael Piotrowski. 2009. A target-driven evaluation of morphological components for German. In Simon Clematide, Manfred Klenner & Martin Volk (Hrsg.), *Searching answers – Festschrift in honour of Michael Hess on the occasion of his 60th birthday*, 85–99. Münster: MV-Wissenschaft.

Piirainen, Elisabeth. 2006. Phraseologie in arealen Bezügen: ein Problemaufriss. In *Linguistik online* 27(2). 195–218.

Piotrowski, Tadeusz. 2009. Die lexikographische Darstellung von Neologismen. In Erika Worbs (Hg.), *Neue Zeiten – neue Wörter – neue Wörterbücher. Beiträge zur Neologismenlexikographie und -lexikologie*, 107–125. Frankfurt a. M.: Peter Lang.

Quasthoff, Uwe, Fabian Schmidt & Erla Hallsteinsdóttir. 2010. Häufigkeit und Struktur von Phraseologismen am Beispiel verschiedener Web-Korpora. In Stefaniya Ptashnyk, Erla Hallsteinsdóttir & Noah Bubendorfer (Hrsg.), *Korpora, Web und Datenbanken: Computergestützte Methoden in der modernen Phraseologie und Lexikographie*, 37–54. Nürnberg: Schneider Verlag Hohengehren.

Stein, Stephan. 2007. Mündlichkeit und Schriftlichkeit aus phraseologischer Perspektive. In Harald Burger, Dmitrij Dobrovol'skij, Peter Kühn & Neal R. Norrick (Hrsg.), *Phraseologie. Phraseology*, 220–236. Berlin/New York: Walter de Gruyter.

Steyer, Kathrin. 2004. Kookkurrenz. Korpusmethodik, linguistisches Modell, lexikografische Perspektiven. In Kathrin Steyer (Hrsg.), *Wortverbindungen, mehr oder weniger fest*, 87–116. Berlin/New York: Walter de Gruyter.